

4th International Symposium of Transport Simulation-ISTS'14, 1-4 June 2014, Corsica, France

Reducing gridlock probabilities via simulation-based signal control

Carolina Osorio^{a*}, Xiao Chen^b, Michael Marsico^c, Mohamad Talas^c, Jingqin Gao^c, Shitao Zhang^c

^a*Massachusetts Institute of Technology*

^b*University of Coimbra*

^c*New York City Department of Transportation*

Abstract

This paper studies a fixed-time signal control problem for a highly congested urban network with multimodal traffic, numerous signalized intersections, short links and a grid-type topology. The design of signal plans that indeed improve traffic conditions for a network with such complex traffic dynamics is a real challenge. In this paper, we propose a simulation-based approach. We use a simulation-based optimization algorithm to identify a signal plan for an area in eastern Manhattan (New York City, USA), where spillbacks frequently occur and impact the flows on major arterials as well as on the access/ egress to the highly congested Queensboro Bridge. We consider a signal control problem where the objective function explicitly considers queue-length information. We compare the performance of the proposed signal plan to that of the prevailing signal plan in the field. The proposed plan indeed improves traffic conditions as measured by a variety of performance metrics.

© 2015 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Selection and/or peer-review under responsibility of the Organizing Committee of ISTS'14

Keywords: queue-length; spillback; simulation-based optimization, signal control, highly congested urban networks

1. Introduction

For uncongested networks, there is no significant queue formulation, but for congested networks, demand typically approaches or even exceeds capacity, leading to the build-up of vehicular queues. Thus, the design of signal plans may differ for uncongested versus congested scenarios. This paper focuses on the design of traffic signal plans for congested networks with complex traffic dynamics. In particular, we consider urban networks with, high levels of congestion, multimodal traffic, numerous signalized intersections, short links and a grid-type topology.

Short links allow for congestion to spatially propagate quickly, leading to complex between-link dependencies. For instance, a link that spills back will impact the performance of its upstream links. The grid-type topology leads to high-dimensional route alternatives, and may lead to complex behavior of travelers as they react to the formation and propagation of congestion. Congested network with both short links and grid-type topologies are highly prone to the occurrence of spillbacks. The spatial propagation of congestion in the form of vehicular spillbacks may have major impacts in the vicinity of major arterials or on/off ramps of freeways. This is the case of the network studied in this paper.

Past work that focus on signal control while considering the occurrence, dynamics and impacts of spillbacks and gridlocks, include [1, 2, 3, 4]. Detailed reviews of queue management in the context of signal control are given, for instance, in [5] and in [6].

In this work, we consider a stochastic simulation-based approach to the design of signal plans for highly congested, multimodal urban networks with grid-type topologies. Given the complexity of traffic dynamics and the complexity of demand-supply interactions in such networks (e.g., driver behavior, numerous transit travel options, pedestrian behavior), the use of detailed traffic simulators has received much attention for both the design and the evaluation of various traffic management strategies.

2. Methodology

2.1. Simulation-based optimization framework

In this work, we consider the use of a calibrated microscopic traffic simulator. These simulators embed the most detailed demand and supply models. They represent individual vehicles, individual travellers, and describe how each of these travellers make decisions (e.g., travel mode, departure time, route, lane-changing). They also embed detailed supply models (e.g., detailed representation of signal plans, of public transport priorities). We use the simulator to study a signal control problem that mitigates the occurrence, length and duration of vehicular-queues and spillbacks.

A general simulation-based signal control problem can be formulated as follows:

$$\min_{x \in \Omega} f(x) = E[F(x, y; p)], \quad (1)$$

where the decision vector x represents the signal control variables (e.g., green times), and the objective function is the expected function of a stochastic network performance metric F (e.g., link speeds, trip travel time), which depends on x as well as on other endogenous simulation variables y (e.g., link flow capacities, route choice probabilities) and exogenous (i.e., fixed) simulation parameters p (e.g., dynamic origin-destination matrices, network topology, transit network). The feasible region Ω is typically a set of analytical differentiable constraints and bound constraints. The main challenge in addressing such a problem is that there is no closed-form expression available for $f(x)$, it can only be estimated via stochastic simulation. Additionally, an accurate estimation involves running numerous simulation replications and is hence computationally costly to obtain.

Traditional simulation-based optimization (SO) algorithms are general-purpose optimization algorithms that treat the microscopic simulator as a black-box, using no structural information about the underlying transportation problem. Hence, they require a large number of simulation runs in order to identify signal plans with improved performance. They are computationally inefficient techniques, and are of little interest for practitioners.

In this paper, we use a simulation-based optimization (SO) algorithm that is computationally efficient. We use the recently proposed algorithm [7], which is designed for continuous generally constrained urban transportation problems. It is an SO metamodel approach that achieves computational efficiency by combining information from the simulator with information from a macroscopic analytical and differentiable traffic model. The latter provides analytical structural information to the SO algorithm. The macroscopic model used is based on finite (space) capacity queueing network theory. It provides a probabilistic description of network performance. For details

regarding the queueing model, see [8]. Details regarding its formulation for an urban network are given in Chapter 4 of [9]. This queueing model describes urban spillbacks via the queueing notion of blocking. It yields an analytical approximation of lane spillback probabilities.

In this paper, we address a signal control problem for an area in eastern Manhattan (New York City, USA), where spillbacks frequently occur and impact the flows on major arterials as well as on the access/ egress to the highly congested Queensboro Bridge. We address a signal control problem with an objective function that considers explicit queue information (e.g., queue length). We address this problem with the SO algorithm mentioned above, and compare the proposed signal plans to the existing signal plan for the area. This paper illustrates with a challenging case study how the SO algorithm in [7] can be used to design traffic management strategies that improve traditional network wide traffic metrics (e.g. travel times) as well as more intricate metrics such as queue information. Ongoing work, to be presented at the conference, compares the performance of various signal problem formulations that use different metrics to account for queue-lengths and spillbacks.

Let us first describe the main ideas underlying the SO algorithm used in this paper. For details regarding its formulation, we refer the reader to [7]. The method has been used to successfully address complex constrained simulation-based signal control problems in a computationally efficient manner [10, 11, 12, 13].

This SO method is a metamodel method. A metamodel is an analytical approximation of the objective function f (as defined in Equation (1)). The main ideas of metamodel SO methods are illustrated in Figure 1.

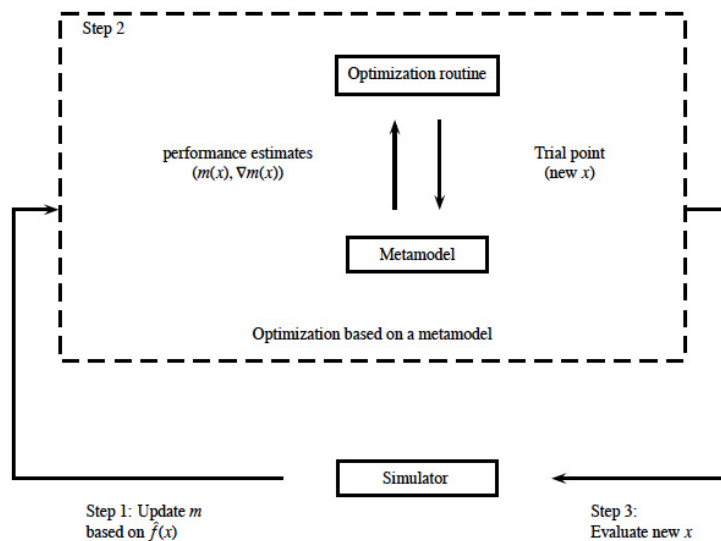


Fig. 1. Metamodel simulation-based optimization methods. Adapted from Alexandrov et al. (1999).

At a given iteration k , the SO algorithm iterates over the following steps: 1) fit the metamodel, m_k , based on the set of simulation observations collected so far, 2) use m_k to perform optimization and derive a trial point x_k , 3) evaluate the performance of this trial point with the simulator, which leads to new simulation observations. As new simulation observations become available, the accuracy of the metamodel can be improved (Step 1), leading to trial points with improved performance (Step 2). These steps are iterated until, for instance, the computational budget is depleted. By resorting to a metamodel approach, the stochastic response of the simulation is replaced by a deterministic metamodel response function, m , such that efficient deterministic optimization techniques can be used. Recent reviews of metamodels are given in [15], [16] and [17]. The algorithm in [7] uses a metamodel that combines a functional and a physical component and has the following functional form:

$$m(x, y; \alpha, \beta, q) = \alpha_0 f_A(x, y; q) + \phi(x; \beta), \quad (2)$$

where ϕ denotes the functional component, f_A (the physical component) represents the approximation of the objective function as derived by an analytical macroscopic traffic model, z are endogenous macroscopic model variables (e.g., queue-length distributions), q are exogenous macroscopic parameters (e.g., total demand). The vector α denotes the set of metamodel parameters. These are fitted by solving a weighted least squares problem that minimizes the distance between the simulation observations and the metamodel predictions. The functional component ϕ is a quadratic polynomial in x with diagonal second-derivative matrix. The physical component f_A is the approximation of the objective function f (Equation (1)) provided by an analytical macroscopic queueing-theoretic traffic model.

The role of this physical component is to provide a problem-specific approximation of the objective function f , unlike the functional component ϕ , which has the same functional form regardless of the objective function chosen. The metamodel can be interpreted as a linear combination of an analytical global and problem-specific approximation of the objective function and a quadratic error term.

2.2. Queue management signal control problem

To illustrate the need for explicitly accounting for queue-lengths in the formulation of the signal control problem, we present a few details regarding the queue-lengths of the network of interest. The study area of interest is a Manhattan sub network that consists of over 130 roads and over 40 intersections. We consider part of the morning peak-period 8am-9am. There are over 11000 vehicle trips for that hour. The queueing network representation of the road network maps the roads into a set of 284 queues. Figure 2 displays for each queue in the network its spillback probability under the signal plan currently used in the field. These probabilities are calculated as follows. We run 50 replications of the simulation model. For each replication and each queue, every three seconds we evaluate the vehicular queue-length. We use these queue-length measurements to estimate over the 8am-9am hour the proportion of time where spillback occurred. This proportion is obtained as an average over both the 8am-9am period of interest and over the 50 simulation replications. These proportions are used as estimates of the spillback probabilities. They are displayed in Figure 2.

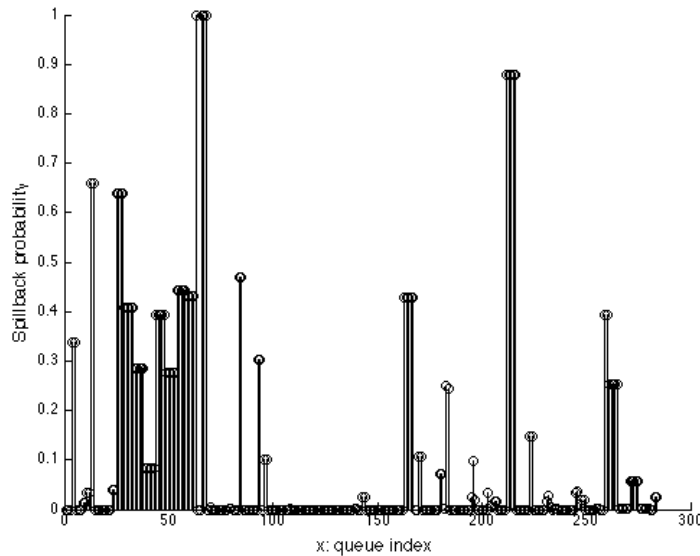


Fig. 2. Spillback probability for each queue under the current signal plan.

This figure illustrates that there are various queues where spillback happens more than 50% of the time. More importantly, even for the queues where the spillback probability is low, the occurrence of spillback may have a significant effect on congestion propagation upstream. This motivates the use of a signal control formulation that explicitly accounts for queue-length metrics.

In order to formulate the signal control problem, we introduce the following notation:

b_i	available cycle ratio of intersection i ;
N_l	queue-length of link l ;
$x(j)$	green split of phase j ;
x_L	vector of minimal green splits;
L	set of links within the area of interest;
I	set of intersection indices;
$P_I(i)$	set of phase indices of intersection i ;

They should also be separated from the surrounding text by one space.

$$\min_x f(x) = \sum_{l \in L} E[N_l(x, y; p)], \quad (3)$$

subject to

$$\sum_{j \in P_I(i)} x(j) = b_i, \quad i \in I \quad (4)$$

$$x \geq x_L. \quad (5)$$

This problem is a fixed-time signal control problem, where the decision variables x are the green splits. In this problem, the stage structure is given, the ϕ sets, the cycle times and the all-red durations are fixed. The performance metric used, $\sum_{l \in L} E[N_l(x, y; p)]$, is the sum of expected queue-lengths. Constraints (4) guarantee that for a given intersection the available cycle time is distributed across all endogenous phases. Constraints (5) ensure lower bounds for the green splits. These are set to 7 seconds, and are based on current New York City Department of Transportation (NYCDOT) practices.

2.3. Physical component

Recall that the metamodel formulation of Equation (2) requires an analytical expression for f_A , which is the approximation of the objective function f as derived by the analytical queueing-theoretic model. Here, we present the analytical and differentiable expression for f_A .

Let Q denote the set of queues that represent the links, L . Then, the objective function can be rewritten as a function of queue metrics, rather than link metrics:

$$\sum_{l \in L} E[N_l(x, y; p)] = \sum_{i \in Q} E[N_i(x, y; p)] \quad (6)$$

We now present how an analytical expression for the expected queue-length of a queue, $E[N_i]$, is derived. We use the analytical queueing-theoretic traffic model derived in Chapter 4 of [9]. We present its formulation below. For further details on its formulation, we refer the reader to [9].

γ_i	external arrival rate
λ_i	total arrival rate
μ_i	service rate
$\tilde{\mu}_i$	unblocking rate

$\hat{\mu}_i$	effective service rate
ρ_i	traffic intensity
P_i^f	probability of being blocked at queue i after service
ℓ_i	space capacity
N_i	number of vehicles in queue i
$P(N_i = \ell_i)$	probability of queue i being full
p_{ij}	transition probability from queue i to queue j
\mathcal{D}_i	set of downstream queues of queue i

$$\lambda_i = \gamma_i + \frac{\sum_j p_{ji} \lambda_j P(N_j < \ell_j)}{P(N_i < \ell_i)} \quad (7)$$

$$\frac{1}{\bar{\mu}_i} = \frac{1}{\mu_i} + P_i^f \frac{1}{\bar{\mu}_i} \quad (8)$$

$$\frac{1}{\bar{\mu}_i} = \sum_{j \in \mathcal{D}_i} \frac{\lambda_j P(N_j < \ell_j)}{\lambda_i P(N_i < \ell_i) \bar{\mu}_j} \quad (9)$$

$$P(N_i = \ell_i) = \frac{1 - \rho_i}{1 - \rho_i^{\ell_i+1}} \rho_i^{\ell_i} \quad (10)$$

$$P_i^f = \sum_j p_{ij} P(N_j = \ell_j) \quad (11)$$

$$\rho_i = \sum_{j \in \mathcal{D}_i} \frac{\lambda_j}{\bar{\mu}_i} \quad (12)$$

In order to approximate the objective function f (of Equation (3)), we proceed as follows. For a given queue i , its expected queue-length is defined as:

$$E[N_i] = \sum_{n=0}^{\ell_i} n P(N_i = n) \quad (13)$$

The stationary marginal queue-length probabilities $P(N_i = n)$ are obtained when evaluating the traffic model, they are given by Equation (10). Combining ideas from Equations (13) and (10), we can obtain the following closed-form expression for $E[N_i]$:

$$E[N_i] = \rho_i \left(\frac{1}{1 - \rho_i} - (\ell_i + 1) \frac{\rho_i^{\ell_i}}{1 - \rho_i^{\ell_i+1}} \right) \quad (14)$$

Hence, the analytical approximation of the objective function is given by:

$$\sum_{i \in Q} E[N_i] = \sum_{i \in Q} \rho_i \left(\frac{1}{1 - \rho_i} - (\ell_i + 1) \frac{\rho_i^{\ell_i}}{1 - \rho_i^{\ell_i+1}} \right) \quad (15)$$

3. Case study

In this case study we use a microscopic traffic simulation model, provided by the NYCDOT, and calibrated one hour of the morning peak period (8am-9am). It is implemented with the simulation software Aimsun [18]. The considered eastern Manhattan network is displayed in Figure 3.

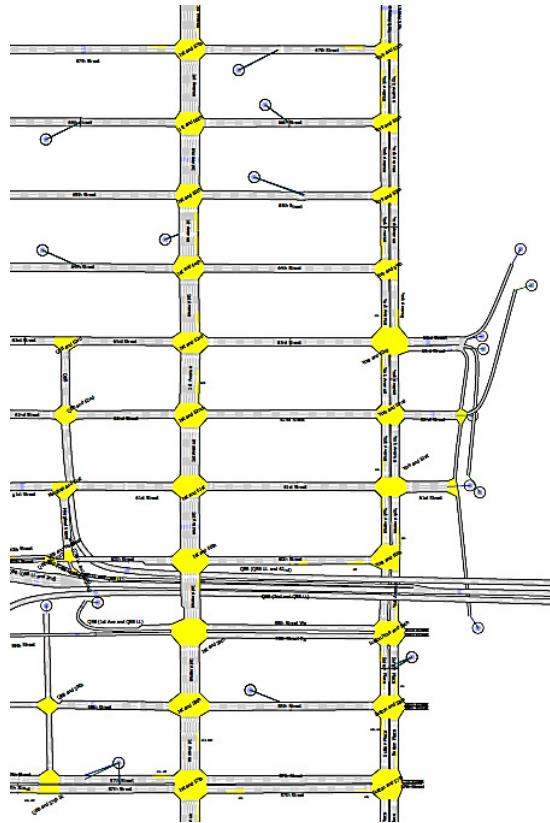


Fig. 3. Aimsun microsimulation network for the upper east Manhattan area.

We compare the performance of the signal plan proposed by the metamodel SO algorithm (denoted m) with the signal plan currently used in the field by NYCDOT for that area. The SO algorithm is run with a computational budget of 150 simulation, i.e., we allow for a total of 150 simulation calls, and stop the algorithm once this

maximum number of simulation runs is reached.

For each signal plan (that obtained by the SO method and the existing one), we evaluate its performance using the simulator. We embed the signal plan within the simulator. We run 50 simulation replications of the simulator, and estimate the following performance metrics: (i) expected total network queue length (this is the sum of the expected queue-length over all queues in the network, it correspond to the SO objective function); (ii) expected trip travel time (minutes); (iii) expected network throughput (veh/h); (iv) expected density in the entire network (veh/km); total spillback probability (this is the sum of the spillback probabilities of all queues).

In order to estimate any of these performance measures for a given signal plan, we run 50 simulation replications. We then plot the empirical cumulative distribution function (cdf) of each performance measure, i.e., each cdf curve consists of 50 simulation estimates of the corresponding performance measure. Each plot of Figure 4 considers a given performance measure, and displays two cdf curves: one for the proposed signal plan (dashed curve denoted “*m*”) and one for the prevailing signal plan (solid curve denoted “existing”).

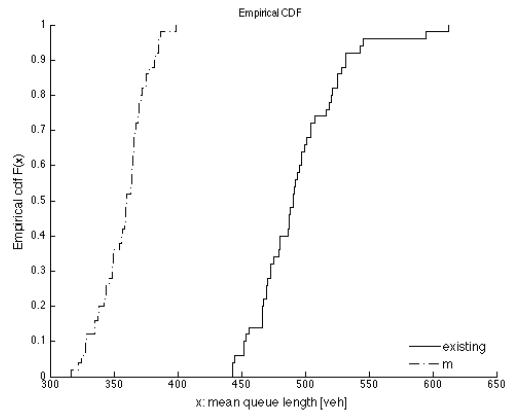
Figure 4(a) shows that the proposed plan leads to a significant reduction in the expected total network queue-length, the average reduction is of the order of 26%. Additionally, for the SO-based plan the highest value observed (among the 50 simulation replications) for the average total network queue-length is of the order of 400 vehicles, which is smaller than the smallest value observed under the existing signal plan. Figure 4(b) shows that the proposed plan reduces the expected trip travel time, with an average reduction of the order of 10%. Figure 4(c) shows that the proposed plan also leads to a reduction in the average network density. Figure 4(e) considers the expected network throughput. The proposed plan increases throughput.

Of particular interest is Figure 4(d), which shows that the total spillback probability is significantly reduced. The average reduction is of the order of 25%. This illustrates that the current signal control formulation clearly achieves the proposed goal of mitigating spillbacks across the network. Additionally, there is much less variability across replications in the total spillback probability for the proposed plan than for the existing plan.

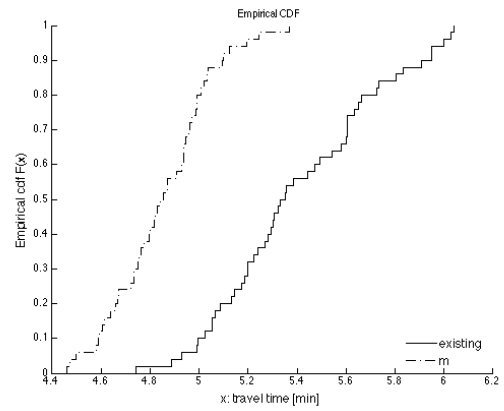
4. Conclusion

In this paper, we address a signal control problem for a highly-congested area in eastern Manhattan (New York City, USA), where spillbacks frequently occur. The network has complex traffic dynamics due to its multimodal congested traffic, short links, numerous signalized intersections and grid-type topology. For such networks it is a great challenge to design signal plans that mitigate the spatial and temporal propagation of congestion.

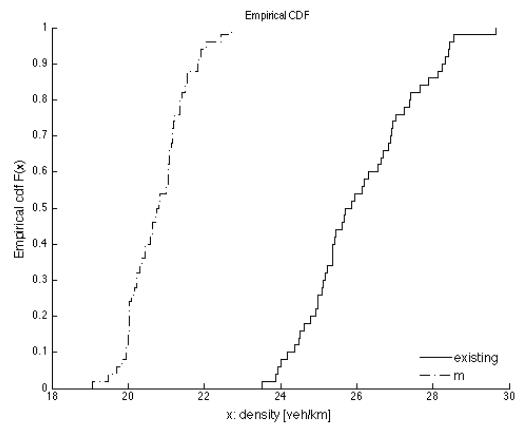
This paper addresses a simulation-based signal control problem, where the objective function explicitly accounts for queue-length information. The aim is to mitigate the occurrence of spillbacks, and hence limit the spatial propagation of congestion. The performance of the proposed plan is compared to that of the existing plan for that area. The proposed plan yields significant improvements when evaluated with various performance metrics. As part of ongoing work, we are studying other problem formulations that account for queue-length and spillback information with different metrics.



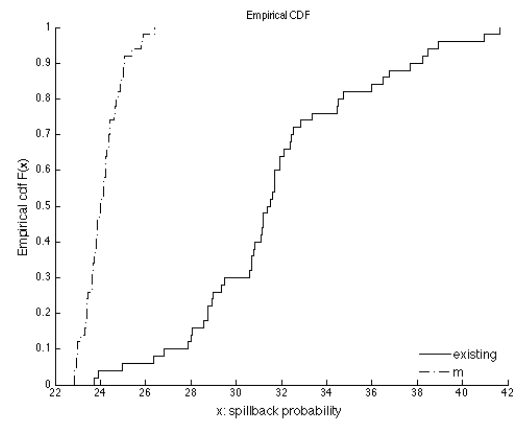
(a) expected total network queue-length



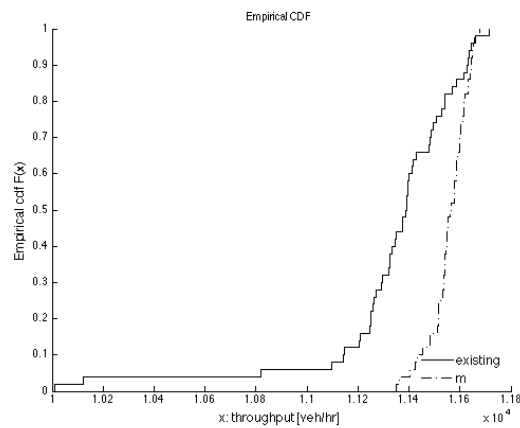
(b) expected trip travel time



(c) expected network density



(d) total spillback probability



(e) expected network throughput

Fig.4. Comparison of the performance of the SO-based signal plan and the current signal plan.

Acknowledgements

The second author of this paper was partially supported by the framework of Project EMSURE - Energy and Mobility for Sustainable Regions (CENTRO-07-0224-FEDER-002004).

References

- [1] Y. Liu, G.-L. Chang, An arterial signal optimization model for intersections experiencing queue spillback and lane blockage, *Transportation research part C: emerging technologies* 19 (1) (2011) 130–144.
- [2] K. Aboudolas, M. Papageorgiou, A. Kouvelas, E. Kosmatopoulos, A rolling-horizon quadratic-programming approach to the signal control problem in large-scale congested urban road networks, *Transportation Research Part C: Emerging Technologies* 18 (5) (2010) 680 – 694. doi:10.1016/j.trc.2009.06.003.
- [3] G. Abu-Lebdeh, R. Benekohal, Development of traffic control and queue management procedures for oversaturated arterials, *Transportation Research Record* 1603 (1997) 119–127.
- [4] P. G. Michalopoulos, G. Stephanopoulos, Oversaturated signal systems with queue length constraintsii: Systems of intersections, *Transportation Research* 11 (6) (1977) 423–428.
- [5] A. Hajbabaie, J. C. Medina, R. F. Benekohal, C. NEXTRANS, Traffic signal coordination and queue management in oversaturated intersections, Tech. rep., Purdue University Discovery Park (2011).
- [6] D. Quinn, A review of queue management strategies, *Traffic Engineering+ Control* 33 (11) (1992) 600–5.
- [7] C. Osorio, M. Bierlaire, A simulation-based optimization framework for urban transportation problems, *Operations Research* 61 (6) (2013) 1333–1345.
- [8] C. Osorio, M. Bierlaire, An analytic finite capacity queueing network model capturing the propagation of congestion and blocking, *European Journal of Operational Research* 196 (3) (2009) 996–1007.
- [9] C. Osorio, Mitigating network congestion: analytical models, optimization methods and their applications, Ph.D. thesis, Ecole Polytechnique Federale de Lausanne (2010).
- [10] C. Osorio, L. Chong, A computationally efficient simulation-based optimization algorithm for large-scale urban transportation, *Transportation Science* Forthcoming.
- [11] C. Osorio, K. Nanduri, Energy-efficient urban traffic management: a microscopic simulation-based approach, *Transportation Science* Forthcoming.
- [12] C. Osorio, K. Nanduri, Emissions mitigation: coupling microscopic emissions and urban traffic models for signal control Submitted.
- [13] X. Chen, C. Osorio, B. F. Santos, A simulation-based approach to reliable signal control, in: *Proceedings of the International Symposium on Transportation Network Reliability (INSTR)*, 2012.
- [14] N. M. Alexandrov, R. M. Lewis, C. R. Gumbert, L. L. Green, P. A. Newman, Optimization with variable-fidelity models applied to wing design, Tech. Rep. CR-1999-209826, NASA Langley Research Center, Hampton, VA, USA (December 1999).
- [15] A. R. Conn, K. Scheinberg, L. N. Vicente, Introduction to derivative-free optimization, MPS/SIAM Series on Optimization, Society for Industrial and Applied Mathematics and Mathematical Programming Society, Philadelphia, PA, USA, 2009.
- [16] R. R. Barton, M. Meckesheimer, Metamodel-based simulation optimization, in: S. G. Henderson, B. L. Nelson (Eds.), *Handbooks in operations research and management science: Simulation*, Vol. 13, Elsevier, Amsterdam, 2006, Ch. 18, pp. 535–574.
- [17] J. Søndergaard, Optimization using surrogate models - by the Space Mapping technique, Ph.D. thesis, Technical University of Denmark (2003).
- [18] TSS, AIMSUN 7 Microsimulator Users Manual, Transport Simulation Systems (2011).